

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: August 22, 2003, 14:09:12 ; Search time 33 Seconds

(without alignments)
4426.210 Million cell updates/sec

Title: US-09-745-506-74

Perfect score: 506

Sequence: 1 GTGATGTTACTTGTGTCCT.....TCTGTACTTAACATTCOA 1553

Scoring table:

ORIGO	Xgapop 60.0	Xgapext 60.0
Xgapop 60.0	Xgapext 60.0	
Fgapop 6.0	Fgapext 7.0	
Delop 6.0	Delext 7.0	

Searched: 127863 seqs, 47026705 residues

Word size: 1

Total number of hits satisfying chosen parameters: 255550

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-O=/cgn2.1/USPRO/spool/US09745506/runat_22082003_132737_11173/app.query.fasta.1.1735
-DB=SwissProt_41 -QFMT=fastan -SUFFIX=n2poli.rsp -MINMATCH=0.1 -LOOPEL=0
-LOOEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR.SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=eto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09745506.ecgn.1.1.40.etrnat.22082003_132737_11173 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG.SCORES=0 -WAIT -DSPELOCK=100 -LONGLOC
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	350	69.2	350	1 NF3L_HUMAN	09q8t8 homo sapien
2	51	10.1	350	1 NF3L_MOUSE	09q8t8 mus musculu
3	10	2.0	366	1 YF59_STAM	099ct7 staphylococ
4	9	1.8	288	1 NIF3_YEAST	P53081 saccharomyc
5	8	1.6	169	1 E1BS_ADECG	065424 canine aden
6	8	1.6	169	1 E1BS_ADECG	P35984 canine aden
7	8	1.6	169	1 E1BS_ADECG	016678 canine aden
8	8	1.6	367	1 CCR3_MOUSE	088410 mus musculu
9	8	1.6	368	1 CCR3_HUMAN	P49682 homo sapien
10	8	1.6	655	1 M1A1_MOUSE	P45700 mus musculu
11	8	1.6	1440	1 POLG_JAEVN	P14403 j genome po
12	8	1.6	1490	1 CDR4_CANAL	074676 candida alb
13	8	1.6	1664	1 RPAL_YEAST	P10964 saccharomyc
14	8	1.6	1689	1 RPAL_SCHRO	P15398 schizosach
15	8	1.6	1780	1 POLG_JAEV	P05769 m genome po
16	8	1.6	3432	1 POLG_JAEV	P19110 j genome po
17	7	1.4	81	1 UCRO_HUMAN	014949 homo sapien
18	7	1.4	81	1 UCRO_MOUSE	09c969 mus musculu

19	7	1.4	119	1 RNPA_MYCPA	091719 mycobacteri
20	7	1.4	122	1 PC32_GIAPU	P22371 claviceps p
21	7	1.4	134	1 VG32_HSV1	000100 ictalurid h
22	7	1.4	137	1 ERV4_YEAST	P53173 saccharomyc
23	7	1.4	146	1 RL28_TRYCR	000822 trypanosoma
24	7	1.4	146	1 SSBS_XENLA	P09380 xenopus lae
25	7	1.4	148	1 SSBR_XENLA	P09381 xenopus lae
26	7	1.4	150	1 COXA_HUMAN	P20674 homo sapien
27	7	1.4	152	1 MGSA_PASMO	09c9n3 pasteurella
28	7	1.4	154	1 Y05B_CAEEL	009254 caenorhabdi
29	7	1.4	165	1 CGHB_HUMAN	P01233 homo sapien
30	7	1.4	165	1 PSAP_SYNY3	P29256 synecocyst
31	7	1.4	173	1 NUGM_ONCMY	P48177 oncorhynch
32	7	1.4	173	1 NUGM_SALSA	092zm2 salmo salar
33	7	1.4	177	1 IL10_SHEEP	029408 ovis aries
34	7	1.4	188	1 AMPK_METFE	P22524 methanother
35	7	1.4	197	1 DCTR_RHOCA	P37740 rhodobacter
36	7	1.4	198	1 NGAL_RAT	P30152 rattus norv
37	7	1.4	198	1 SDC4_MOUSE	035988 mus musculu
38	7	1.4	201	1 NUGM_CVACA	P48925 cyanidium c
39	7	1.4	211	1 IPYR_SOLITU	043187 solanum tub
40	7	1.4	218	1 Y279_MYCGE	P47521 mycoplasma
41	7	1.4	221	1 IFE3_ARATH	09f459 arabidopsis
42	7	1.4	221	1 PSB1_ORYSA	064464 oryza sativ
43	7	1.4	224	1 FLGA_YEREN	056892 yersinia en
44	7	1.4	224	1 MYOG_HUMAN	P15173 homo sapien
45	7	1.4	224	1 MYOG_MOUSE	P12979 mus musculu

ALIGNMENTS

RESULT 1

NF3L_HUMAN STANDARD; PRT; 350 AA.

AC 09q8t8; 09h2d2; 09hcl8;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE NIF3-like protein 1 (Amphotropic lateral sclerosis 2 chromosomal

DE region candidate gene protein 1) (My018 protein) (MDS015).

GN NIF3L1 OR ALS2CR1.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI_TaxId=9606;

RN [1]

RP SEQUENCE FROM N.A., AND SUBCELLULAR LOCATION.

RX MEDLINE=20573864; PubMed=11124544;

RA Tascou S., Uedelhoven J., Dixkens C., Neyerula K., Engel W.,

RA Burfeind P.;

RT "Isolation and characterization of a novel human gene, NIF3L1, and its

RT mouse ortholog, Nif3l1, highly conserved from bacteria to mammals.;"

RL Cytogetnet. Cell Genet. 90:330-336(2000).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=21100893; PubMed=11161814;

RA Hadano S., Yanagisawa Y., Skaug J., Fichter K., Nasir J.,

RA Martindale D., Koop B.F., Scherer S.W., Nicholson D.W., Rouleau G.A.,

RA Ikeda J.-E., Hayden M.R.;

RT "Cloning and characterization of three novel genes, ALS2CR1, ALS2CR2,

RT and ALS2CR3, in the juvenile amyotrophic lateral sclerosis (ALS2)

RT critical region at chromosome 2q33-q34: candidate genes for ALS2.;"

RL Genomics 71:200-213(2001).

RN [3]

RP SEQUENCE FROM N.A.

RC TISSUE=Brain;

RA Mao Y.M., Xie Y., Huang X.Y., Ying K., Dai J.L.;

RA Submitted (Apr-1998) to the EMBL/Genbank/DDJ databases.

RN [4]

RP SEQUENCE FROM N.A.

RA Huang C., Qian B., Tu Y., Gu W., Wang Y., Han Z., Chen Z.;

RT "Novel genes expressed in hematopoietic stem/progenitor cells from

RT myelodysplastic syndromes patient.;"

RL submitted (SEP-1999) to the EMBL/Genbank/DBJ databases.
[5]
RP SEQUENCE FROM N.A.
RC TISSUE-Ovarian carcinoma;
RA Itogi T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Matsukawa M., Hosokita T., Kaku Y., Kodaira H., Kondo H.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuho Y.,
RA Niromiya K., Iwayanagi T.,
RT "NDO human cDNA sequencing project."
RL submitted (AUG-2000) to the EMBL/Genbank/DBJ databases.
[6]
RP SEQUENCE FROM N.A.
RC TISSUE-Skin;
RX MEDLINE-22389257; PubMed-12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Ditschenko L., Maruska K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uscdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length
RL human and mouse cDNA sequences."
CC Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE UPF0135 (NIF3) FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: AF283538; AAC44846.1; ALT_INIT.
DR EMBL: AB038949; BAB32499.1; -
DR EMBL: AF060513; AAG43131.1; -
DR EMBL: AF182416; AAG14952.1; -
DR EMBL: AK023378; BAB14551.1; -
DR EMBL: BC007654; AAH07654.1; ALT_INIT.
DR Genew; HGNC:13390; NIF3L1.
DR Interpro; IPR002678; DUF34.
DR Pfam; PF01784; NIF3; 1.
DR TIGRFAMs; TIGR00486; TIGR00486; 1.
FT CONFLICT 77 TWNTWK -> NLEHMR (IN REF. 4).
SQ SEQUENCE 350 AA; 38983 MW; 81FA5AD35B25ED7 CRC64;

Alignment Scores:
Pred. No.: 0 Length: 350
Score: 350.00 Matches: 350
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 69.17% Indels: 0
DB: 1 Gaps: 0

US-09-745-506-74 (1-1553) x NF3L_HUMAN (1-350)
QY 245 ATGGATTGAAGGCTCTCCCTTCTCCTTGAATGATGATTCGATCCGCTGTTGCGAG 304
|||||

DB 1 MetaspLeuLysAlaLeuLeuSerLeuLysnAspPheAlaSerLeuSerPheAlaGlu 20
QY 305 AGTTGGACAAATGTTGATTAATCTGGTGGACCAACCCACACATATGTAATACATC 364
|||||
DB 21 SerTrpAspAsnValGlyLeuLeuValGluProSerProPronHisThrValAsnThrLeu 40
QY 365 TTTCGGACCAATGACCTGACGTAGAGAACTGATGAGAGAGGTCTGCAAAAGAGCAGAC 424
|||||
DB 41 PheLeuThrAsnAspLeuThrGluGluValMetGluGluValLeuGlnLysLysAlaAsp 60
QY 425 CTCATCTCTCTACATCCGCTATCTCTCCGACCCAGAACGATACGATACCTGGACACCA 484
|||||
DB 61 LeuLeuLeuSerTyrHisProProThrLeuPheMetGlySerGlyLeuThrPheAsnThr 80
QY 485 TGAAGAGACGCGCTGGTATCCGGGCTCTGGAGAACAGATCGGATATCTATCTCTCAT 544
|||||
DB 81 TrpLysGluArgLeuValIleArgAlaLeuGlnAsnArgValGlyIleTyrSerProHis 100
QY 545 ACAGCCTATGATGCTGGCCGCCAGGCGCTCAACAACATGCTGGCTAAAGGCTGGAGCT 604
|||||
DB 101 ThrAlaTyrAspAlaAlaProGlnGlyValAsnAsnTrpLeuAlaLysGlyLeuGlyAla 120
QY 605 TGTACCTCCAGGCCCATCATCTCCCAAGCTCCCAACCTCCATCCCTACAGAGGAAACAC 664
|||||
DB 121 CysThrSerArgProIleHisProSerTyrAspProAsnTyrProThrGluLysnHis 140
QY 665 CGAGTAGAATCAACGTTAACTACACCCCAAGACCTGCAACAAGTCATGCTGCAAGTGA 724
|||||
DB 141 ArgValGluProAsnValAsnTyrThrGlnAspLeuAsnLysValMetSerAlaValLys 160
QY 725 GGAATGACGGTGTCTGCTGATCTTTCTTTCTGCTAGAGCTGTAATGAGAAACA 784
|||||
DB 161 GlyIleAspGlyValSerValThrSerPheSerAlaArgThrGlyAsnGluGlnThr 180
QY 785 CGGATTTATCTGAATTTGACTCAGAAAGCTTTGATGCAAGGTGAGATTTCTTCCCG 844
|||||
DB 181 ArgIleAsnLeuAsnCysThrGlnLysAlaLeuMetGlnValAlaAspPheLeuSerArg 200
QY 845 AACAAACAACTTTATCGAAGAACAGCAAAATCTGCTACGTGAGAACCTTTGCTTACAT 904
|||||
DB 201 AsnLysGlnLeuTyrGlnLysThrGlnLeuSerLeuGlnLysProLeuLeuLys 220
QY 905 ACTGGATGGACGGTATATGACACCTGGATGATCTGCTCTCCGCAACCATGATGAT 964
|||||
DB 221 ThrGlyMetGlyArgLeuCysThrLeuAspLysSerValSerLeuAlaThrMetIleAsp 240
QY 965 CGAATTAACAAACACCTAAACATATCTCATCTTACGCTTGAAGCTTGGGGTGGGGAAC 1024
|||||
DB 241 ArgIleLysArgHisLeuLysLeuSerHisIleArgLeuAlaLeuGlyValArgThr 260
QY 1025 TTGAGTCTCAAGTCAAAAGTGTGGCCCTGTGCTGCTGCTGGGAGCAGCGTTCGAG 1084
|||||
DB 261 LeuGlnSerGlnValLysValAlaIleLeuCysAlaGlySerValLeuGln 280
QY 1085 GGTGTGAGGCTGCTTACCTGACAGAGTGAAGTGCATGATGATGATGATGATGATGAT 1144
|||||
DB 281 GlyValIleAlaAspLeuTyrLeuThrGlyLeuMetSerHisIleAspThrLeuAspAla 300
QY 1145 GCTTCCCAAGAAATTAATGTCATCTCTGTCGACACAGCAACATGAAAGAGGCTTCTT 1204
|||||
DB 301 AlaSerGlnGlyIleAsnValIleLeuCysGlnHisSerAsnThrGluArgPheLeu 320
QY 1205 TCTGACCTTGAGATATGCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1264
|||||
DB 321 SerAspLeuArgAspMetLeuAspSerHisLeuGlnLysnLysIleAsnIleLeuSer 340
QY 1265 GAGAGCTACAGGAGACCTCTTCAGCTGTA 1294
|||||
DB 341 GluThrAspArgAspProLeuGlnValVal 350

RESULT 2
NF3L_MOUSE
ID NF3L_MOUSE STANDARD; PRT; 350 AA.

AC Q9EQ80; Q9D098;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE NIF3-like protein 1.
 GN NIF3L1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
 RX MEDLINE=2057364; PubMed=1124544;
 RA Tascou S., Uedelhoven J., Dixkens C., Nayeria K., Engel W.,
 RA Burfelind P.;
 RT "Isolation and characterization of a novel human gene, NIF3L1, and its
 RT mouse ortholog, Nif3l1, highly conserved from bacteria to mammals.";
 RL Cytogenet. Cell genet. 90:330-336(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Embryo;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schiml L.M., Staubl F., Suzuki R., Tomita M., Wagner L., Mashio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli P., Boujunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Kombarits P.,
 RA Norioka P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wyszynski B.A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection";
 RL Nature 409:685-690(2001).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- TISSUE SPECIFICITY: Ubiquitous.
 CC -1- SIMILARITY: BELONGS TO THE UPF0135 (NIF3) FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AF284439; BAA45961.1; ALT_INIT.
 DR EMBL; AK011670; BAB27769.1; ALT_INIT.
 DR MGD; MGI:1929485; NIF3L1.
 DR InterPro: IPR002678; DUF34.
 DR Pfam: PF01784; NIF3; 1.
 DR TrEMBL; TIGR00486; TIGR00486; 1.
 FT CONFLICT 164 164 G -> S (IN REF. 2).
 FT CONFLICT 178 178 E -> K (IN REF. 2).
 FT CONFLICT 191 191 L -> F (IN REF. 2).
 FT CONFLICT 195 195 L -> O (IN REF. 2).
 FT CONFLICT 198 198 L -> F (IN REF. 2).
 FT CONFLICT 208 208 T -> I (IN REF. 2).
 SQ SEQUENCE 350 AA; 38828 MW; B0FA71503FCF086 CRC64;

Alignment Scores:

Pred. No.: 5 71e-44 Length: 350
 Score: 51.00 Matches: 51
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0

Query Match: 10.08% Indels: 0
 DB: 1 Gaps: 0
 US-09-745-506-74 (1-1553) x NIF3L_MOUSE (1-350)
 QY 245 ATGATTTGAGAGCGCTCTCTCTCTCTCTGTAATGACTTCCTGCTGTTGCTGAG 304
 DB 1 MetaspLeuLyAlaLeuLeuSerSerLeuAaAnspPheAlaSerLeuSerPheAlaGlu 20
 QY 305 AGTTGGACAAATGTTGGATTACTGTGGACCAAGCCACACATCTCTAAATACAC 364
 DB 21 SerrTPasPasValGlyLeuLeuValGluProSerProProHisThrValAsnThrLeu 40
 QY 365 TTCCTGACCAATGACTGCTGAGGAGGATG 397
 DB 41 PheLeuThrAsnAspLeuThrGluValMet 51
 RESULT 3
 YF59_STAM STANDARD; PRT; 366 AA.
 ID YF59_STAM
 AC 099TT7;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DE Hypothetical protein SAV1559/SA1388.
 GN SAV1559 OR SA1388
 OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and
 OS Staphylococcus aureus (strain N315).
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
 OX NCBI_TaxID=158878, 158879;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Mu50 / ATCC 700699, and N315;
 RX MEDLINE=21311952; PubMed=11418146;
 RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
 RA Kai M., Oguchi A., Aoki K.-I., Nagai Y., Iino T., Ito T.,
 RA Kanamori M., Matsunaru H., Maruyama A., Murakami H., Hosoyama A.,
 RA Mizutani-Ul Y., Takahashi N.K., Sawano T., Inoue R.-I., Katto C.,
 RA Sekimizu K., Hirakawa H., Kohara S., Goto S., Yabuuchi J.,
 RA Kakenisha M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
 RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
 RT "Whole genome sequencing of methicillin-resistant Staphylococcus
 RT aureus";
 RL Lancet 357:1225-1240(2001).
 CC -1- SIMILARITY: BELONGS TO THE UPF0135 (NIF3) FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AP003362; BAB57721.1; -;
 DR EMBL; AP003134; BAB42651.1; -;
 DR PIR: F89936; F89936;
 DR InterPro: IPR002678; DUF34.
 DR Pfam: PF01784; NIF3; 1.
 DR TrEMBL; TIGR00486; TIGR00486; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 366 AA; 41075 MW; CA7D5FADF078FF4 CRC64;

Alignment Scores:

Pred. No.: 0.158 Length: 366
 Score: 10.00 Matches: 10
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.98% Indels: 0
 DB: 1 Gaps: 0

US-09-745-506-74 (1-1553) x YF59_STAM (1-366)

QY 239 GCTGAGAGTGGACATGTTGATTACG 328
| | | | |
Db 19 AlaGUserTTPASPASValcIyleuLeu 28

RESULT 4
NIF3_YEAST STANDARD: PRT; 288 AA.
ID NIF3_YEAST
AC P53081;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE NGG1-interacting factor 3.
GN NIF3 OR YGL221C.
OS Saccharomyces cerevisiae (Baker's yeast).
CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c;
RX MEDLINE=97435481; PubMed=9290212;
RA Rieger M., Brueckner M., Schaefer M., Mueller-Auer S.;
RT "Sequence analysis of 203 kilobases from Saccharomyces cerevisiae
chromosome VII."
RL Yeast 13:1077-1090(1997).
CC -1 SIMILARITY: BELONGS TO THE UPF0135 (NIF3) FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
DR EMBL: Z72743; CAA96937.1; -
DR PIR: S64243; S64243.
DR SGD: S0003189; NIF3.
DR InterPro: IPR002678; DUF34.
DR Pfam: PF01784; NIF3; 1.
DR TIGRfam: TIGR00486; TIGR00486; 1.
DR SEQUENCE 288 AA; 31888 MW; B6AB6E4BAFA4776A CRC64;
SQ

Alignment Scores:
Pred. No.: 1.77 Length: 288
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.78% Indels: 0
Gaps: 0
DB: 1

US-09-745-506-74 (1-1553) x NIF3_YEAST (1-288)
QY 1178 CACAGCAGACCTGACGAGCTTCTT 1204
| | | | |
Db 252 HisSeranThrgluArglyPheLeu 260

RESULT 5
EIBS_ADECC STANDARD: PRT; 169 AA.
ID EIBS_ADECC
AC Q65942;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE EIB protein, small T-antigen (Early EIB 19 kDa protein).
DE Canine adenovirus type 1 (strain CL1).
OS Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
CC
CC NCBI_TaxID=69150;
RN [1]
RP SEQUENCE FROM N.A.
RA Campbell J.B., Zhao Y.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
DR EMBL: U55001; AAB05430.1; -
DR InterPro: IPR002924; AdenoEIB_19kDa.
DR InterPro: IPR002475; BCL2 family.
DR Pfam: PF01691; Adeno_EIB_19k; 1.
DR ProDom: PD004074; AdenoEIB_19kDa; 1.
DR PROSITE: PS50062; BCL2_FAMILY; 1.
KM Early protein.
SQ SEQUENCE 169 AA; 19333 MW; F8A4DC68B8D0DFE4 CRC64;
SQ

Alignment Scores:
Pred. No.: 20.6 Length: 169
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.58% Indels: 0
Gaps: 0
DB: 1

US-09-745-506-74 (1-1553) x EIBS_ADECC (1-169)
QY 1384 CTTCCAGAGAGTGCTTCGAGGT 1407
| | | | |
Db 138 LeuProGluSerValPheGluGly 145

RESULT 6
EIBS_ADECC STANDARD: PRT; 169 AA.
ID EIBS_ADECC
AC P35984;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DE 01-NOV-1997 (Rel. 35, Last annotation update)
DE EIB protein, small T-antigen (Early EIB 19 kDa protein).
OS Canine adenovirus type 1 (strain Glaxo).
CC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
CC NCBI_TaxID=10513;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=90163565; PubMed=2623943;
RA Spibey N., McCleary R.S., Cavanagh R.M.A.;
RT "Identification and nucleotide sequence of the early region 1 from
RT canine adenovirus types 1 and 2."
RL Virus Res. 14:241-256(1989).
DR PIR: C60010; C60010.
DR InterPro: IPR002924; AdenoEIB_19kDa.
DR InterPro: IPR002475; BCL2 family.
DR Pfam: PF01691; Adeno_EIB_19k; 1.
DR ProDom: PD004074; AdenoEIB_19kDa; 1.
DR PROSITE: PS50062; BCL2_FAMILY; 1.
KM Early protein.
SQ SEQUENCE 169 AA; 19364 MW; 4B7F31A74136623F CRC64;
SQ

Alignment Scores:
Pred. No.: 20.6 Length: 169
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.58% Indels: 0
Gaps: 0
DB: 1

US-09-745-506-74 (1-1553) x EIBS_ADECC (1-169)
QY 1384 CTTCCAGAGAGTGCTTCGAGGT 1407
| | | | |
Db 138 LeuProGluSerValPheGluGly 145

RESULT 7

EIBS_ADECR
ID EIBS_ADECR STANDARD: PRT: 169 AA.
AC 096678;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE E1B protein, small T-antigen (Early E1B 19 kDa protein).
OS Canine adenovirus type 1 (strain R1261).
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
OX NCBI_TaxID=69151;
RN [1]
RP MEDLINE=97275900; PubMed=9129661;
RA Morrison M.D., Onions D.E., Nicolson L.;
RT *Complete DNA sequence of canine adenovirus type 1.*;
RL J. Gen. Virol. 78:873-878(1997).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Y07760; CAA69053.1; -
DR InterPro: IPR002924; AdenoE1B_19kDa.
DR InterPro: IPR002475; BCL2_family.
DR Pfam: PF01691; Adeno_E1B_19k; 1.
DR ProDom: PD004074; AdenoE1B_19kDa; 1.
DR ProSITE: PS50062; BCL2_FAMILY; 1.
KW Early protein.
SQ SEQUENCE 169 AA; 19367 MW; F8A4D2868ED0FE4 CRC64;

Alignment Scores:
Pred. No.: 20.6 Length: 169
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.58% Indels: 0
DB: 1 Gaps: 0

US-09-745-506-74 (1-1553) x EIBS_ADECR (1-169)
OY 1384 CTTCCAGAGAGTGTTCGAGGGT 1407
DB 138 LeuProGluSerValPheGluGly 145

RESULT 8
ID CCR3_MOUSE STANDARD: PRT: 367 AA.
AC 088410;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE C-X-C chemokine receptor type 3 (CXCR-3) (CXCR-3).
GN CXCR3 OR CMKAR3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=98318636; PubMed=9653165;
RA Soto H., Wang W., Strieter R.M., Copeland N.G., Gilbert D.J.,
RA Jenkins N.A., Hedrick J., Zlotnik A.;
RT "The CC chemokine 6CKine binds the CXC chemokine receptor CXCR3.";
RT Proc. Natl. Acad. Sci. U.S.A. 95:8205-8210(1998).
CC -1- FUNCTION: RECEPTOR FOR SCYB9/MIG, SCYB10/INP10 AND SCYB11/ITAC (BY
CC SIMILARITY). BINDS TO SCYA21.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF045146; AAC40163.1; -
DR MGI: MGI:1277207; Cxcr3.
DR InterPro: IPR004070; CXC_chemkine3.
DR InterPro: IPR002276; GPCR_Rhodopsn.
DR Pfam: PF00001; 7tm_1; 1.
DR PRINTS: PR01532; CXCCMKINER3.
DR PRINTS: PR00237; GPCRHHODOPSN.
DR PROSITE: PS00237; G-PROTEIN_RECP_FL_1; 1.
DR PROSITE: PS50262; G-PROTEIN_RECP_FL_2; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein.
FT DOMAIN 1 52
FT TRANSMEM 53 79
FT DOMAIN 80 88
FT TRANSMEM 89 109
FT DOMAIN 110 124
FT TRANSMEM 125 146
FT DOMAIN 147 168
FT TRANSMEM 169 188
FT DOMAIN 189 211
FT TRANSMEM 212 232
FT DOMAIN 233 254
FT TRANSMEM 255 276
FT DOMAIN 277 297
FT TRANSMEM 298 320
FT DOMAIN 321 367
FT DISULFID 123 202
FT CARBOHYD 22 22
FT CARBOHYD 32 32
SQ SEQUENCE 367 AA; 41016 MW; 029FBB778E3CD4EA CRC64;

Alignment Scores:
Pred. No.: 18.6 Length: 367
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.58% Indels: 0
DB: 1 Gaps: 0

US-09-745-506-74 (1-1553) x CCR3_MOUSE (1-367)
OY 744 TCACCTCTTTTCTGCTAGACTG 767
DB 60 SerLeuLeuPheLeuGluGlyLeu 67

RESULT 9
ID CCR3_HUMAN STANDARD: PRT: 368 AA.
AC P49682; O15185;
DT 01-FEB-1996 (Rel. 33, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DE C-X-C chemokine receptor type 3 (CXCR-3) (CXCR-3) (CD183
DE antigen).
GN CXCR3 OR GPR9.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=97188912; PubMed=9064356;
RA Itoescher M., Gerber B., Loetscher P., Jones S.A., Piali L.,
RA Clark-Lewis I., Baggiolini M., Moser B.;
RT "Chemokine receptor specific for IP10 and mlg: structure, function,
CC -----

RT and expression in activated T-lymphocytes.";
 RL J. Exp. Med. 184:965-969(1996).
 RN [2]
 RA SEQUENCE FROM N.A.
 RA Gutierrez J., Varona R., Zaballeros A., Lind P., Marquez G.;
 RA submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN, Lung, and Testis;
 RX MEDLINE-22388257; PubMed-12477932;
 RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshlyuk S., Caranci P., Prange C.,
 RA Raha S.S., Loguettano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Vallatton D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko V., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield J.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [4]
 RP SEQUENCE OF 5-368 FROM N.A.
 RX MEDLINE-96115583; PubMed-8666380;
 RA Marchese A., Heiber M., Nguyen T., Heng H.H.Q., Saldivia V.R.,
 RA Cheng R., Murphy P.M., Tsui L.-C., Shi X., Gregor P., George S.R.,
 RA O'Dowd B.F., Docherty J.M.;
 RT "Cloning and chromosomal mapping of three novel genes, GPR9, GPR10,
 RT and GPR14, encoding receptors related to Interleukin 8, neuropeptide
 RL Y, and somatostatin receptors.";
 RL Genomics 29:335-344(1995).
 RN [5]
 RP LIGAND BINDING.
 RC TISSUE-Fetal astrocytes;
 RX MEDLINE-98290735; PubMed-9625760;
 RA Cole K.E., Strick C.A., Paradis T.J., Ogborne K.T., Loetscher M.,
 RA Glade R.P., Lin W., Boyd J.G., Moser B., Wood D.E., Sahagan B.G.,
 RA Neote K.;
 RT "Interferon-inducible T cell alpha chemottractant (I-TAC): A novel
 RT non-ELR CXC chemokine with potent activity on activated T cells
 RT through selective high affinity binding to CXCR3.";
 RL J. Exp. Med. 187:2009-2021(1998).
 CC -1- FUNCTION: RECEPTOR FOR SCYB9/MIG, SCYB10/INP1 AND SCYB11/ITAC.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC -1- DATABASE: NAME=PROX; NOTE=PROX 1:78-83(2000).
 CC WWW="http://www.ncbi.nlm.nih.gov/prox/guide/650534941_g.htm".
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X95876; CAAG5126.1; ALT. INT.
 DR EMBL: 279783; CAB02143.1; ALT. INT.
 DR EMBL: BC034403; AAH34403.1;
 DR EMBL: U32674; AAC30505.1;
 DR Genew: HGNC:4540; CXCR3.
 DR MIM: 600894;
 DR GO: GO:0005737; C:cytoplasm; TAS.
 DR GO: GO:0005887; C:Integral to plasma membrane; TAS.

DR GO: GO:0006960; P:antimicrobial humoral response (sensu Inver. . .); TAS.
 DR GO: GO:0007155; P:cell adhesion; TAS.
 DR GO: GO:0006928; P:cell motility; TAS.
 DR GO: GO:0006935; P:chemotaxis; TAS.
 DR GO: GO:0007204; P:cytosolic calcium ion concentration elevation; TAS.
 DR InterPro: IPR004070; KXC_chemkine3.
 DR InterPro: IPR000276; GPCR_Rhodopsn.
 DR Pfam: PF00001; 7tm_1; 1.
 DR PRINTS: PR01532; CXCCHKINER3.
 DR PRINTS: PR00237; GPCR_RHODOPSIN.
 DR PROSITE: PS00237; G-PROTEIN_RECPT_FL_1; 1.
 DR PROSITE: PS0262; G-PROTEIN_RECPT_FL_2; 1.
 KW G-protein coupled receptor; Transmembrane; Glycoprotein; Antigen.
 FT DOMAIN 1
 FT TRANSMM 54
 FT DOMAIN 81
 FT TRANSMM 90
 FT DOMAIN 111
 FT TRANSMM 125
 FT DOMAIN 148
 FT TRANSMM 170
 FT DOMAIN 190
 FT TRANSMM 213
 FT DOMAIN 234
 FT TRANSMM 256
 FT DOMAIN 278
 FT TRANSMM 299
 FT DOMAIN 322
 FT DISULFID 124
 FT CARBOHYD 22
 FT CARBOHYD 32
 FT CONFLICT 75
 SQ SEQUENCE 368 AA; 40659 MW; F08A3B44B2BAD04 CRC64;
 Alignment Scores:
 Pred. No.: 18.6
 Score: 8.00
 Percent Similarity: 100.00%
 Best Local Similarity: 100.00%
 Query Match: 1.58%
 DB: 1
 Gaps: 0
 Length: 368
 Matches: 8
 Conservative: 0
 Mismatches: 0
 Indels: 0
 US-09-745-506-74 (1-1553) x CCR3_HUMAN (1-368)
 QY 744 TCACCTCTTTTCTGCTAGAGACTG 767
 DB 61 SerleuLeuPheleuLeuGlyleu 68
 RESULT 10
 M1AL_MOUSE STANDARD; PRT; 655 AA.
 ID M1AL_MOUSE
 AC P45700;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Manno5Y1-oligosaccharide 1,2-alpha-mannosidase IA (EC 3.2.1.113)
 DE (Processing alpha-1,2-mannosidase IA) (Alpha-1,2-mannosidase IA)
 DE (Mannosidase alpha class IA member 1) (Man(9)-alpha-mannosidase).
 GN MAN1A1 OR MAN1A.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BALB/C;
 RX MEDLINE-94193679; PubMed-8144580;
 RA Lal A., Schutzbach J.S., Forsee W.T., Neame P.J., Moremen K.W.;
 RT "Isolation and expression of murine and rabbit cDNAs encoding an
 RT alpha 1,2-mannosidase involved in the processing of asparagine-linked
 RT oligosaccharides.";
 RL J. Biol. Chem. 269:9872-9881(1994).
 CC -1- FUNCTION: Involved in the maturation of Asn-linked

CC oligosaccharides. Progressively trim alpha-1,2-linked mannose
CC residues from Man(9)GlcNAc(2) to produce Man(5)GlcNAc(2).
CC -1- CATALYTIC ACTIVITY: Hydrolysis of the terminal 1,2-linked alpha-D-
CC mannose residues in the oligo-mannose oligosaccharide
CC Man(9)(GlcNAc)(2).
CC -1- COFACTOR: Calcium.
CC -1- ENZYME REGULATION: Inhibited by both 1-deoxymannojirimycin and
CC kifunensine (By similarity).
CC -1- PATHWAY: N-GLYCOSYLATION.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Golgi.
CC -1- SIMILARITY: BELONGS TO FAMILY 47 OF GLYCOSYL HYDROLASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U04299; AAA17747.1; -
CC PIR: A54408; A54408.
CC
CC HSP: P32906; IDL2.
CC MGD: MGI:104677; Man1a.
CC GO: GO:0000139; C:Golgi membrane; IDA.
CC InterPro: IPR001382; Glyco_hydro_47.
CC Pfam: PF01532; Glyco_hydro_47; 1.
CC PRINTS: PR00747; GLYHDRLASE47.
CC DR PRODOM: PD003239; Glyco_hydro_47; 1.
CC KW Hydrolyase; Glycosidase; Glycoprotein; Transmembrane; Calcium-binding;
CC Signal-anchor; Golgi stack.
CC FT DOMAIN 1 43 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 44 64 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT CA_BIND 65 655 LUMENAL (POTENTIAL).
FT CARBOHYD 271 282 POTENTIAL.
FT N_LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 655 AA; 73375 MW; 15AF658364930BFA CRC64;

Alignment Scores:
Pred. No.: 17.2 Length: 655
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.58% Indels: 0
Gaps: 0

US-09-745-506-74 (1-1553) x M1A1_MOUSE (1-655)
QY 487 GAAGAGCGCGCTGTGATCCGGCG 510
|||||
Db 109 GUGUGYALAPROGLYASPPOGLY 116

RESULT 11
POLG_JAEVN STANDARD; PRT; 1440 AA.
ID POLG_JAEVN
AC P14403; P08769; 13, Created
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polypeptide [Contains: Capsid protein C (Core protein); Matrix
DE protein (Envelope protein M); Major envelope protein E; Nonstructural
DE proteins NS1, NS2A, and NS2B; Protease/helicase (EC 3.4.21.98) (NS3)]
DE (Fragment).
OS Japanese encephalitis virus (strain Nakayama).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Flavivirus.
OX NCBI_TaxID=11076;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=67236200; PubMed=3035787;
RA McCa P.C., Mason P.W., Schmaljohn C.S., Dalrymple J.M., Mason T.L.,
RA Fournier M.J.;

RT "Partial nucleotide sequence of the Japanese encephalitis virus
RT genome.";
RL . Virology 158:348-360(1987).
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polypeptide, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND RNA.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M16574; AAA46251.1; -
CC PIR: A27844; GNMVJF.
CC HSP: P14336; 15VB.
CC InterPro: IPR001122; Flav1_capsidC.
CC InterPro: IPR000336; Flav1_glycoprote.
CC InterPro: IPR000069; Flav1_M.
CC InterPro: IPR001157; Flav1_NS1.
CC InterPro: IPR000752; Flav1_NS2A.
CC InterPro: IPR000487; Flav1_NS2B.
CC InterPro: IPR002535; Flav1_propep.
CC Pfam: PF01003; Flav1_capsid; 1.
CC Pfam: PF02832; Flav1_glycop_C; 1.
CC Pfam: PF00869; Flav1_glycoprot; 1.
CC Pfam: PF01004; Flav1_M; 1.
CC Pfam: PF00948; Flav1_NS1; 1.
CC Pfam: PF01005; Flav1_NS2A; 1.
CC Pfam: PF01002; Flav1_NS2B; 1.
CC Pfam: PF01570; Flav1_propep; 1.
CC PRODOM: PD001566; Flav1_glycoprote; 1.
CC DR Polypeptide; Glycoprotein; Core protein; Coat protein;
CC Envelope protein; Hydrolyase; Helicase; ATP-binding; Transmembrane;
CC Nonstructural protein 1.
CC KW NON_TER 1
FT CHAIN 1 53 CAPSID PROTEIN C.
FT PROPEP 54 146 ENVELOPE GLYCOPROTEIN M.
FT CHAIN 147 222 MAJOR ENVELOPE PROTEIN E.
FT CHAIN 223 794 NONSTRUCTURAL PROTEIN NS1.
FT CHAIN 795 1136 NONSTRUCTURAL PROTEIN NS2A.
FT CHAIN 1137 1301 NONSTRUCTURAL PROTEIN NS2B.
FT CHAIN 1302 1432 PROTEASE/HELICASE (NS3).
FT CHAIN 1433 >1440
FT DISULFID 225 252 BY SIMILARITY.
FT DISULFID 282 338 BY SIMILARITY.
FT DISULFID 296 327 BY SIMILARITY.
FT DISULFID 314 343 BY SIMILARITY.
FT DISULFID 412 509 BY SIMILARITY.
FT DISULFID 526 557 BY SIMILARITY.
FT CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 376 376 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 852 852 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 929 929 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON_TER 1440 1440
SQ SEQUENCE 1440 AA; 158184 MW; 4DA89A365A3C2B6E CRC64;

Alignment Scores:
Pred. No.: 15.5 Length: 1440
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.58% Indels: 0

DB: 1 Gaps: 0

US-09-745-506-74 (1-1553) x POLG_JAEVN (1-1440)

QY 1056 GTGCTGCTCTGGAGACACCTTC 1079
 |||||||
 DB 1180 ValLeuValLeuGlyAlaAlaPhe 1187

RESULT 12

CDR4_CANAL STANDARD: PRT; 1490 AA.

AC 074676;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE ABC transporter CDR4.
 GN CDR4.
 OS Candida albicans (Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
 OX NCBI_TaxId=5476;
 [1]
 RN RP SEQUENCE FROM N.A.
 RC STRAIN=1161;
 RX MEDLINE=98440373; PubMed=9767132;
 RA Franz R., Michel S., Morschauser J.,
 RT "A fourth gene from the Candida albicans CDR family of ABC
 RT transporters.";
 RL Gene 220:91-98(1998).
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. PD55 SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF044921; AAC72295.1; -
 DR PIR; T30550; T30550.
 DR InterPro: IPR003439; AAA_ATPase.
 DR InterPro: IPR003439; ABC_transporter.
 DR InterPro: IPR005285; PDR.
 DR Pfam; PF00005; ABC_tran; 2.
 DR ProDom; PD000006; ABC_transporter; 2.
 DR SMART; SM00382; AAA; 2.
 DR TIGRFAMs; TIGR00956; 3a01205; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
 DR PROSITE; PS00893; ABC_TRANSPORTER_2; 2.
 KW ATP-binding; Transmembrane; Glycoprotein; Transport; Repeat.
 FT DOMAIN 1 516
 FT TRANSMEM 517 537 POTENTIAL.
 FT TRANSMEM 551 571 POTENTIAL.
 FT TRANSMEM 601 621 POTENTIAL.
 FT TRANSMEM 626 646 POTENTIAL.
 FT TRANSMEM 659 679 POTENTIAL.
 FT TRANSMEM 767 787 POTENTIAL.
 FT DOMAIN 788 1182 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 1183 1203 POTENTIAL.
 FT TRANSMEM 1217 1237 POTENTIAL.
 FT TRANSMEM 1268 1288 POTENTIAL.
 FT TRANSMEM 1304 1324 POTENTIAL.
 FT TRANSMEM 1333 1353 POTENTIAL.
 FT TRANSMEM 1455 1475 POTENTIAL.
 FT NP BIND 882 889 ATP (POTENTIAL).
 SO SEQUENCE 1490 AA; 169341 MW; 858676C872A4F5C5 CRC64;

Alignment Scores:

Pred. No.: 15.4 Length: 1490
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0

Query Match: 1.61% Indels: 0

DB: 1 Gaps: 0

US-09-745-506-74 (1-1553) x CDR4_CANAL (1-1490)

QY 1028 CTAGGTTCTCCACCCCAAGG 1005
 |||||||
 DB 1395 LeuArgPheSerProProGingly 1402

RESULT 13

RPA1_YEAST STANDARD: PRT; 1664 AA.

AC P10964; Q99330;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE DNA-directed RNA polymerase I 190 kDa polypeptide (EC 2.7.7.6) (A190).
 GN RPA1 OR RPA190 OR RRN1 OR YOR341W OR O6276.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxId=4932;
 [1]
 RN RP SEQUENCE FROM N.A.
 RC MEDLINE=88139333; PubMed=2830265;
 RA Memet S., Gouy M., Marck C., Sentenac A., Buhler J.-M.,
 RT "RPA190, the gene coding for the largest subunit of yeast RNA
 RT polymerase A.";
 RL J. Biol. Chem. 263:2830-2839(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=5288c / FY73;
 RX MEDLINE=97103776; PubMed=8948102;
 RA Purnelle B., Goffeau A.;
 RT "Nucleotide sequence analysis of a 40 kb segment on the right arm of
 RT yeast chromosome XY reveals 18 open reading frames including a new
 RT private kinase and three homologues to chromosome I genes.";
 RL Yeast 12:1475-1481(1996).
 CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
 CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
 CC SUBSTRATES. RNA POLYMERASE I IS ESSENTIALLY USED TO TRANSCRIBE
 CC RIBOSOMAL DNA UNITS.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
 CC (RNA)(N).
 CC -1- SUBUNIT: RNA POLYMERASE I CONSISTS OF 14 DIFFERENT SUBUNITS. THIS
 CC SUBUNIT IS THE LARGEST COMPONENT OF RNA POLYMERASE I.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES WERE
 CC FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA
 CC PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE
 CC III FOR 5S AND 5.8S GENES.
 CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; J03530; AAA34890.1; -
 DR EMBL; X95720; CAA65029.1; -
 DR EMBL; 275249; CAA99665.1; -
 DR PIR; S67250; S67250.
 DR SGD; S0005868; RPA190.
 DR InterPro: IPR000722; RNA_pol_A.
 DR InterPro: IPR007080; RNA_pol_Rp1_1.
 DR InterPro: IPR007066; RNA_pol_Rp1_3.
 DR InterPro: IPR007083; RNA_pol_Rp1_4.
 DR InterPro: IPR007081; RNA_pol_Rp1_5.
 DR InterPro: IPR006593; RNA_pol_A_N.
 DR Pfam; PF04997; RNA_pol_Rp1_1; 1.

DR Pfam: PF00623; RNA_pol_Rpb1_2; 1.
 DR Pfam: PF04983; RNA_pol_Rpb1_3; 1.
 DR Pfam: PF05000; RNA_pol_Rpb1_4; 1.
 DR Pfam: PF04998; RNA_pol_Rpb1_5; 1.
 DR SMART; SM00663; RPOA_N; 1.
 KW Transferrase; DNA-directed RNA polymerase; Transcription; zinc;
 FT zinc-finger; Nuclear protein.
 FT ZN-FING 62 78 C2H2-TYPE (POTENTIAL).
 FT CONFLICT 158 158 N -> T (IN REF. 1).
 SQ SEQUENCE 1664 AA; 186431 MW; DF65A7AA459D5E6D CRC64;

Alignment Scores:
 Pred. No.: 15.2 Length: 1664
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.61% Indels: 0
 DB: 1 Gaps: 0

US-09-745-506-74 (1-1553) x RPA1_YEAST (1-1664)

OY 591 TTAGCCACGAGTGTGACGCC 568
 DB 557 LeuAlaAsnGlnLeuLeuThrPro 564

RESULT 14
 RPA1_SCHPO STANDARD; PRT; 1689 AA.

AC P15398;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE DNA-directed RNA polymerase I 190 kDa polypeptide (EC 2.7.7.6).
 GN RPA1 OR NUCL OR SPBC4C3.05C.
 OS Schizosaccharomyces pombe (fission yeast).
 OC Schizosaccharomycetes; Fungi; Ascomycota; Schizosaccharomycetales;
 OC Schizosaccharomycetes.
 OC NCBI_Taxid=4896;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89139548; PubMed=2537310;
 RA Hirano T., Konoha G., Toda T., Yanagida M.;
 RT "Essential roles of the RNA polymerase I largest subunit and DNA
 RT topoisomerases in the formation of fission yeast nucleolus.";
 RL J. Cell Biol. 108:243-253(1989).
 RN (2)
 RP SEQUENCE FROM N.A.
 RX STRAIN=972 / HM123;
 RX MEDLINE=89232741; PubMed=2854522;
 RA Yamagishi M., Nomura M.;
 RT "Cloning and sequence determination of the gene encoding the largest
 RT subunit of the fission yeast Schizosaccharomycetes pombe RNA polymerase
 RT I.";
 RL Gene 74:503-515(1988).
 RN (3)
 RP SEQUENCE FROM N.A.
 RX STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Spours J., Peat N., Hayes J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeh C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volckart G., Bert R., Robben J., Gymnopoulos B.,
 RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,

RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
 RA Beer P., Zimmermann W., Medler H., Wambutt R., Pinnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaire V., Motter S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt K., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Galliard C., Tallade V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Carrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shipkovski G.V., Usery D., Barrell B.G., Nurse P.;
 RT "The genome sequence of Schizosaccharomyces pombe.";
 RL Nature 415:871-880(2002).
 CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
 CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
 CC SUBSTRATES. RNA POLYMERASE I IS ESSENTIALLY USED TO TRANSCRIBE
 CC RIBOSOMAL DNA UNITS.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC (RNA)(N).
 CC -1- SUBUNIT: EACH CLASS OF RNA POLYMERASE IS ASSEMBLED FROM 9 TO 14
 CC DIFFERENT POLYPEPTIDES. THIS SUBUNIT IS THE LARGEST COMPONENT OF
 CC RNA POLYMERASE I.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES WERE
 CC FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA
 CC PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE
 CC III FOR 5S AND 5.8S GENES.
 CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.

 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: X14783; CAJ32887.1; -;
 DR EMBL: M37411; AAA35326.1; -;
 DR EMBL: A1021730; CAA16827.1; -;
 DR PIR: JS0080; JS0080.
 DR Genedb_Spombe: SPBC4C3.05C; -;
 DR InterPro: IPR000722; RNA_pol_A.
 DR InterPro: IPR007080; RNA_pol_Rpb1_1.
 DR InterPro: IPR007066; RNA_pol_Rpb1_3.
 DR InterPro: IPR007083; RNA_pol_Rpb1_4.
 DR InterPro: IPR007081; RNA_pol_Rpb1_5.
 DR InterPro: IPR006592; RNA_pol_A_N.
 DR Pfam: PF04997; RNA_pol_Rpb1_1; 1.
 DR Pfam: PF00623; RNA_pol_Rpb1_2; 1.
 DR Pfam: PF04983; RNA_pol_Rpb1_3; 1.
 DR Pfam: PF05000; RNA_pol_Rpb1_4; 1.
 DR Pfam: PF04998; RNA_pol_Rpb1_5; 1.
 DR SMART; SM00663; RPOA_N; 1.
 KW Transferrase; DNA-directed RNA polymerase; Transcription; zinc;
 KW zinc-finger; Nuclear protein.
 FT ZN-FING 63 79 C2H2-TYPE (POTENTIAL).
 FT CONFLICT 69 69 D -> A (IN REF. 1).
 FT CONFLICT 84 84 I -> S (IN REF. 1).
 FT CONFLICT 704 704 T -> I (IN REF. 1).
 FT CONFLICT 1581 1581 A -> T (IN REF. 1).
 FT CONFLICT 1681 1681 T -> N (IN REF. 1).
 SQ SEQUENCE 1689 AA; 189244 MW; 2D2D3A2DEC94A497 CRC64;

Alignment Scores:
 Pred. No.: 15.2 Length: 1689
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.61% Indels: 0
 DB: 1 Gaps: 0

US-09-745-506-74 (1-1553) x RPA1_SCHPO (1-1689)

QY 591 TTAGCACAACGATTGTTGACGCC 568
 |||||
 Db 565 LeuAlaAsnGlnLeuLeuThrPro 572

RESULT 15
 POLG_MVEV STANDARD; PRT: 1780 AA.
 AC P05769;
 DT 01-NOV-1988 (Rel. 09, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polyprotein (Contains: Capsid protein C (Core protein); Matrix protein (Envelope protein M); Major envelope protein E; Nonstructural proteins NS1, NS2A, and NS2B; Protease/helicase (PC 3.4.21.98) (NS3)) (Fragment).
 OS Murray valley encephalitis virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Flavivirus.
 CX NCBI_TaxID=11079;
 RN [1]
 RX SEQUENCE FROM N.A.
 RA MEDLINE=66200215; PubMed=3009829;
 RA Dalgarno L., Trent D.W., Strauss J.R., Rice C.M.;
 RT "Partial nucleotide sequence of the Murray Valley encephalitis virus genome. Comparison of the encoded polypeptides with yellow fever virus structural and non-structural proteins.";
 RL J. Mol. Biol. 187:309-323(1986).
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the P6 position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sdb.ch/announce/> or send an email to license@sdb-sdb.ch).
 CC
 CC EMBL: X03467; CAA27184.1;
 CC PIR: A24635; GNMVW.
 CC HSSP: P14336; 1SVB.
 CC MEROPS: S07.001;
 DR InterPro: IPR001122; Flav1_capsidC.
 DR InterPro: IPR000336; Flav1_glycoprote.
 DR InterPro: IPR001850; Flav1_helicase.
 DR InterPro: IPR000069; Flav1_M.
 DR InterPro: IPR001157; Flav1_NS1.
 DR InterPro: IPR000752; Flav1_NS2A.
 DR InterPro: IPR000487; Flav1_NS2B.
 DR InterPro: IPR002535; Flav1_propep.
 DR Pfam: PF01003; Flav1_capsid.1.
 DR Pfam: PF02832; Flav1_glycop.1.
 DR Pfam: PF00869; Flav1_glycoprot.1.
 DR Pfam: PF00949; Flav1_helicase.1.
 DR Pfam: PF01004; Flav1_M.1.
 DR Pfam: PF00948; Flav1_NS1.1.
 DR Pfam: PF01005; Flav1_NS2A.1.
 DR Pfam: PF01002; Flav1_NS2B.1.
 DR Pfam: PF01570; Flav1_propep.1.
 DR ProDom: PD001556; Flav1_glycoprote.1.
 DR ProDom: PD001496; Flav1_NS1.1.
 KW Polyprotein; Glycoprotein; Core protein; Coat protein;
 KW Envelope protein; Hydrolyase; Helicase; ATP-binding; Transmembrane;
 KW Nonstructural protein.
 FT INIT_MET 1 REMOVED FROM CAPSID PROTEIN C BY THE

FT CHAIN 1 121
 FT PROPEP 122 217
 FT CHAIN 218 292
 FT CHAIN 293 793
 FT CHAIN 794 1207
 FT CHAIN 1208 1372
 FT CHAIN 1373 1503
 FT CHAIN 1504 >1780
 FT CHAIN 44 60
 FT TRANSMEM 112 128
 FT TRANSMEM 278 294
 FT TRANSMEM 773 789
 FT TRANSMEM 1178 1194
 FT TRANSMEM 1219 1235
 FT TRANSMEM 1250 1266
 FT TRANSMEM 1312 1328
 FT TRANSMEM 1378 1394
 FT TRANSMEM 1401 1417
 FT TRANSMEM 1476 1492
 FT TRANSMEM 295 322
 FT DISULFID 352 408
 FT DISULFID 366 397
 FT DISULFID 384 413
 FT DISULFID 482 580
 FT DISULFID 597 628
 FT CARBOHYD 73 73
 FT CARBOHYD 140 140
 FT CARBOHYD 446 446
 FT CARBOHYD 923 923
 FT CARBOHYD 968 968
 FT CARBOHYD 1000 1000
 FT NON_TER 1780 1780
 SQ SEQUENCE 1780 AA; 194866 MW; 0D6AA7FF0EB706DE CRC64;

Alignment Scores:
 Pred. No.: 15.1 Length: 1780
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.58% Indels: 0
 DB: 1 Gaps: 0

US-09-745-506-74 (1-1553) x POLG_MVEV (1-1780)

QY 1056 GTGCTGTTCTGGAGACGCTTC 1079
 |||||
 Db 1251 ValLeuValLeuGlyAlaAlaPhe 1258

Search completed: August 22, 2003, 14:33:47
 Job time : 51 secs